Our Ref.: A 1428 PCT
Max-Planck-Gesellschaft zur
Förderung der Wissenschaften e.V. et al.

Nucleic acid molecules encoding proteins which impart the adhesion of Neisseria cells to human cells

The present invention relates to nucleic acid molecules from bacteria of the genus Neisseria encoding proteins mediating the adhesion of Neisseria cells to human cells. Furthermore, the present invention relates to the proteins encoded by these nucleic acid molecules and to antibodies directed against them. The present invention further relates to pharmaceutical compositions, vaccines and diagnostic compositions containing said nucleic acid molecules, proteins and/or antibodies.

To the genus Neisseria (gram-negative cocci) belong a number of bacterial species which, being saprophytes, populate the upper human respiratory tract. Apart from commensal species (e.g.: N. sicca) and opportunistically pathogenic species (e.g.: N. lactamica), two Neisseria species are known which clearly possess human-pathogenic properties. One of the species is N. gonorrhoeae, the pathogen of the venereal disease gonorrhea, which exclusively occurs in humans, and N. meningitidis, the pathogen of the bacterial epidemic meningitis. In both cases the etiology, that is the causal connection between the development of the clinical picture and the population by bacteria from said species has meanwhile been substantiated.

The purulent meningitis (Meningitidis cerebrospinalis epidemica) caused by N. meningitidis ("meningococcus"), which usually is epidemical, is a systemic invasive infection of the human meninx and spinal meninx. Occasionally, hemorrhagic exanthema at the trunk or concomitant diseases caused by Herpes simplex can be observed in addition. The pathogen can appear in the form of several serotypes, which are distinguishable by means of agglutination assays with immune sera. The main groups differ remarkably, and their prevalence differs with regard to when and where they appear. Meningococcus meningitidis has up to now occurred in large numbers every 8 to 12 years with the increased prevalence lasting between 4 to 6 years. While serovar B meningococci brought on 50% to 55% of the recent diseases

to the civilian population as well as to the military personnel in the United States, most epidemic diseases in the United States during the first half of the century were caused by serovar A meningococci.

The clinical picture caused by N. gonorrhoeae usually is an infection localized to the mucous membranes, in most cases of the urogenital tract (gonorrhea), more rarely of the conjunctiva (conjunctivis gonorrhoeae, gonoblennorrhoe), which is acquired by new born children perinatally, by adults usually unilaterally by smear infection. In very rare cases bacteremia and sepsis occur after hematogenous dissemination. As a consequence, exanthema with hemorrhagic pustules, diseases from the rheumatic *Formenkreis*, arthritis gonorrhoica and/or endocarditis can occur.

Usually, the diseases caused by N. gonorrhoeae and N. meningitidis are treated with antibiotics. More and more, however, the bacteria are becoming resistant to single or groups of the antibiotics used so that the therapy method that has nearly exclusively been used up to now will most likely not be successful in the long run. Therefore, it is desirable and urgent that alternative therapy methods, preferably preventive ones, be developed.

Neisseria gonorrhoeae and N. meningitidis exclusively occur in humans. They have adapted to the host organism and show a number of properties that are able to make the defense mechanisms of the host ineffective. Therefore, up to now there is no vaccine available that prevents gonorrhea. This is to a limited extent also true for meningococcus meningitidis. Even though the disease has recently been caused mainly by bacteria of the same serovar, group B, no effective vaccine against meningococci of group B has existed up to now. Vaccines against other serovars only offer partial protection and are not unproblematic from an immunological point of view. The reason for the failure of the immune defense is, inter alia, the antigen variation of the pathogens, which in the case of the pathogenic Neisseria is particularly developed. However, a limitation of the free development of the antigen variation seems to be necessary where the functional region has to be sterically maintained in order to guarantee the interaction with conserved and constant structures of the host receptors. This requirement especially applies to the adhesins that serve for adhering to the host cell. Only if the functional region that is involved in the physical interaction is kept constant, the interaction with the receptor of the host cell is possible. This region should be excluded from antigen variation to a large

extent and is therefore a suitable starting-point for the development of a new therapy method.

The initial phase of infections usually is the stable adhesion of the pathogens to the host tissue. By interactions between structures of the cell surface of the pathogens and the cell surface of the host cell a mechanically stable linkage is formed that allows the bacteria to stay on the tissue of the host (colonization) and to subsequently propagate locally. The adhesion to the host cell can be divided into two phases with different structures being involved in the interaction.

In the first phase of adhesion a contact between host cell and pathogen is mediated. Often cell appendage organelles, the so-called pili, are involved in mediating the contact. These cell organelles, which are also called fimbriae or fibrils, are few to several fine filamentous rigid or flexible appendages of the bacterial cell, which can be several times as long as the cell diameter. Therefore, there is no contact between the cell walls of pathogen and host cell in the pilus mediated adhesion. The majority of the known pili are heteropolymeric structures consisting of several components. The main subunit, which usually is present in many copies, fulfills the structural function, that is the framework function, whereas the actual adhesion function is fulfilled by side components, which usually are present in few copies.

A further form of adherence is the adhesion of pathogens to the host cells without the contribution of pili (pilus independent adherence, pia). In this case, the pathogen and the host cell are approaching each other, and finally the cell walls directly touch. This adhesion and stabilization of the contact between the cells takes place with the contribution of adhesines that are located in the bacterial cell wall. As a result of the direct contact between the cells, a signal is finally transmitted that initiates the pathogen induced phagocytosis and starts the invasion process into the target cell. The pia form of adherence can autonomically effect the adhesion of pathogens, for example in the case of pathogens lacking pili. It can, however, also act as the second phase of adhesion, that is as the consecutive reaction after pilus mediated adhesion, and stabilize the contact between the cells. The adhesines that are involved in the pilus independent adhesion can but do not necessarily have to show different binding specifities from those that are involved in pilus dependent adhesion.

In the context of the invention the bacterial structures that are involved in the adhesion will in the following be called adhesines, those of the host cells will be called receptors. If there is no contact between adhesin and receptor, "defense mechanisms" of the host, such as fibrillation of the epithelia, mucus secretion, mass flow of body fluids and the like, eliminate the pathogens. The development of an infection is, therefore, prevented from the very beginning. Thus, a disturbance of the adhesion of the pathogens by means of inhibiting the interaction between adhesin and receptor of the target cell represents a very effective approach for preventing and treating infections. Such therapeutically effective approaches comprise the production of antibodies specifically blocking the adhesin function, either by active immunization (vaccination) or by administration of antibodies already existing (passive immunization). The adhesin receptor binding can, in the same way, be inhibited by means of passive administration of both receptor analogous and adhesin analogous substances. These substances competitively bind to the corresponding partner structures, thereby blocking their involvement in productive interactions. In the context of the invention such substances are called inhibitors.

The approaches using pilin, the main component of the pilus that fulfills the structural function, in order to develop a broadly effective vaccine effectively blocking the adhesion of pathogenic Neisseria have failed so far. The reason probably is that (i) pilin itself has no adhesin function and (ii) pilin possesses an especially distinct intra- and interstem specific antigenic variation. Since both limitations, as described above, do not apply to adhesins, the use of an adhesin as a vaccine is more promising.

The technical problem of the present invention therefore is to provide proteins and DNA molecules encoding them that serve as adhesion structures for Neisseria species or contribute to the development of such structures.

This problem is solved by providing the embodiments described in the claims.

Therefore, the present invention relates to nucleic acid molecules containing the nucleotide sequence described in Seq ID No. 1 or parts thereof with these nucleic

acid molecules comprising one or more open reading frames encoding proteins or biologically active fragments thereof from bacteria of the genus Neisseria that mediate the adhesion of Neisseria cells to human cells. The term "reading frame" in this context is used synonymously with the term "coding region".

The subject matter of the invention also relates to nucleic acid molecules that basically show the nucleotide sequence described in Seq ID No. 1 but whereby the nucleotide sequences of the open reading frames deviate from those described in Seq ID No. 1 due to the degeneration of the genetic code. Preferably, the open reading frames of those nucleic acid molecules have nucleotide sequences encoding proteins with one of the amino acid sequences described in Seq ID No. 1.

The subject matter of the invention further relates to nucleic acid molecules hybridizing to the nucleic acid molecules described above and comprising coding regions encoding proteins that mediate the adhesion of Neisseria cells to human cells.

In the context of the present invention the term "hybridization" is used as described in Sambrook et al. (Molecular Cloning, A Laboratory Manual; Cold Spring Harbor Laboratory Press (1989), 1.101 to 1.104). Preferably, this term has the meaning of hybridization under stringent conditions. In particular, it has the meaning of a hybridization that still shows a positive hybridization signal after being washed for 1 h with 1 x SSC and 0.1 % SDS, preferably with 0.2 x SSC and 0.1 % SDS, at 55 °C, preferably at 62 °C and most preferably at 68 °C.

In a preferred embodiment the nucleic acid molecule of the invention originates from a pathogenic Neisseria species, in particular from Neisseria gonorrhoea or Neisseria meningitidis.

The term "nucleic acid molecule" as used here according to the invention relates to the polymeric form of nucleotides of any length, either as ribonucleotides or as desoxyribonucleotides. The term only relates to the primary structure of the molecule. In this sense, it comprises DNA and RNA molecules, in single- or double-stranded form. The DNA can either be cDNA or genomic DNA. The term further comprises the non-modified form as well as scientifically known modifications, e.g., methylation, capping, base substitution with natural or synthetic analogues,

internucleotide modifications with uncharged compounds (e.g., methyl phosphate, phosphoamidate, carbamate, phosphotriester and the like) or with charged compounds (e.g., phosphorothioate, phosphorodithioate and the like) or with binding components such as proteins and peptides (e.g., nucleases, toxins, antibodies, poly-L-lysine, and the like). The term also comprises forms with intercalating substances (e.g., acridin, psoralen, and the like), chelators (e.g., with metals, radioactive metals or oxidizing metals and the like), with alkylating agents and finally with modified bonds (e.g., alpha anomeric nucleic acids, and the like).

The invention also relates to vectors containing a nucleic acid molecule of the invention. The vector can be any prokaryotic or eukaryotic vector. Examples of prokaryotic vectors are chromosomal vectors, such as bacteriophages (e.g., bacteriophage lambda, P1), and extrachromosomal vectors, such as plasmids with circular plasmids being particularly preferred. Suitable prokaryotic vectors are, for example, described in Sambrook et al. (see above), chapters 1 to 4. The vector according to the invention can also be a eukaryotic vector, for example a yeast vector or a vector suitable for higher cells (e.g., a plasmid vector, a viral vector, a plant vector, and the like). Examples of such vectors are also described in Sambrook et al. (see above, chapter 16). A vector containing a nucleic acid molecule of the invention is, for example, plasmid pES25 (contained in the E. coli strain H 2560 (DSM 10257)). The E. coli strain H 2560 was deposited on September 18, 1995 with Deutsche Sammlung von Mikroorganismen (DSM) [German collection of microorganisms] in Brunswick, Federal Republic of Germany, as international recognized depositary authority in accordance with the stipulations of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure under accession number DSM 10257.

The invention furthermore relates to host cells containing a vector as described above or being genetically manipulated with a nucleic acid molecule as described above. The term "host cell" in the context of this invention comprises both prokaryotic and eukaryotic host cells. Prokaryotic cells are preferred, particularly gram-negative prokaryotic cells, in particular E. coli cells. Suitable eukaryotic host cells are, for example, fungal cells (e.g., yeast cells), animal or plant cells.

The nucleotide sequence described in Seq ID No. 1 comprises three open reading frames. They represent an operon forming a functional unity. The three open reading frames called orfI, orfA and orfB encode three proteins that in the context of this invention are called OrfI, OrfA and OrfB. These sequences are responsible for the expression of a protein in Neisseria cells, in particular of the protein OrfA, which is involved in the adhesion of Neisseria cells to human cells. The proteins OrfI and OrfB obviously possess a regulatory function or a function as factors that are able to influence the functionality of OrfA.
This nucleic acid molecule therefore represents a region of the Neisseria genome that encodes proteins having the adhesin function of Neisseria cells.

The present invention further relates to nucleic acid molecules encoding a lipoprotein or biologically active fragments thereof from bacteria of the genus Neisseria having the amino acid sequence as described in Seq ID No. 2. In a preferred embodiment the invention relates to nucleic acid molecules encoding a protein having the amino acid sequence from the amino acid residue 19 to the amino acid residue 320 of the amino acid sequence as described in Seq ID No. 2. Such nucleic acid molecules preferably have the nucleotide sequence described in Seq ID No. 2, in particular the nucleotide sequence from nucleotide 189 to nucleotide 1095 of the sequence described in Seq ID No. 2.
The subject matter of the invention also relates to nucleic acid molecules encoding a lipoprotein from bacteria of the genus Neisseria whereby their nucleotide sequence deviates from the nucleic acid molecules described above due to the degeneration of the genetic code.
Furthermore, the present invention relates to nucleic acid molecules encoding a lipoprotein from bacteria of the genus Neisseria and hybridize to one of the nucleic acid molecules described above (for the definition of the term "hybridization" see above).

The subject matter of the invention also relates to fragments, derivatives and allelic variants of the nucleic acid molecules described above that encode the lipoprotein described above. Fragments are understood to be parts of the nucleic acid molecules that are long enough to encode the protein described. The term derivative

in this context means that the nucleotide sequences of these molecules differ at one or more positions from the sequences of the nucleic acid molecules described above and that they show a high level of homology to these nucleotide sequences. Homology means a sequence identity of at least 40 %, in particular an identity of at least 60 %, preferably of more than 80 % and particularly preferred of more than 90 %. The deviations to the nucleic acid molecules described above can be caused by deletion, substitution, insertion or recombination.

Homology further means that there is a functional and/or structural equivalence between the corresponding nucleic acid molecules or the proteins encoded by them. The nucleic acid molecules that are homologous to those described above and that represent derivatives of these nucleic acid molecules usually are variants of these molecules displaying modifications that have the same biological function. They can be naturally occurring variants, for example sequences from other organisms, or mutations, which either occur naturally or that have been introduced by means of specific mutagenesis. Furthermore, the variants can be synthetically produced sequences.

The allelic variants can be both naturally occurring variants or variants that were synthetically produced or that were produced by recombinant DNA techniques.

The proteins encoded by the various variants of the nucleic acid molecules according to the invention show certain common characteristics, for example enzyme activity, molecular weight, immunological reactivity, conformation etc., as well as physical properties such as the electorphoretic mobility, chromatographic behavior, sedimentation coefficients, solubility, spectroscopic properties, stability, pH optimum, temperature optimum etc.

Preferably, the proteins encoded by the nucleic acid molecules according to the invention show a homology of 80 %, particularly preferred of more than 90 % to the nucleotide sequence described in Seq ID No. 2.

The nucleic acid molecules described above encode a lipoprotein from bacteria of the genus Neisseria. This protein is called OrfA in the context of the present invention. This protein is, according to experimental data, located on the cell surface of Neisseria cells, in particular on the outer membrane. The protein preferably has a molecular weight of about 36 kd if it is analyzed in the T7 expression system.

Furthermore, this protein possesses a biological activity that mediates the adhesion of Neisseria cells to human cells. This is in particularly made because this protein forms a complex with the protein from Neisseria known as PilC. The adhesion preferably takes place on human epithelial cells.

Furthermore, the invention relates to vectors containing nucleic acid molecules described above. Examples of such vectors have already been described above.

In a preferred embodiment the DNA molecules according to the invention are linked in such vectors with regulatory DNA elements that make the expression of the protein in prokaryotic or eukaryotic cells possible. Examples thereof are in the context of this invention promoters, operators, enhancers and the like.

Furthermore, the invention relates to host cells that contain vectors according to the invention described above or that have been genetically manipulated with the nucleic acid molecules described above. Genetically manipulated means that such a molecule has been introduced into the host cell or in a precursor cell by means of (gene)technological methods. Again, the above-described host cells are suitable.

The invention also relates to methods for the production of the described lipoprotein or a biologically active fragment thereof whereby the host cells described above are cultivated under conditions that allow the expression of the protein and the protein is isolated from the cells and/or the culture supernatant.

The invention also relates to proteins encoded by one of the nucleic acid molecules described above, as well as to biologically active fragments thereof as well as to proteins available by the method described above. In particular, the invention relates to proteins having amino acid sequences that immunologically cross-react with the described proteins. The term "protein" comprises in the context of the present invention also naturally occurring variants or modifications or fragments or synthetically produced modifications, variants or fragments with the corresponding biological activity. Derived or recombinant proteins do not necessarily have to be biologically translated from the nucleotide sequence. They can be produced in any

way, including chemical synthesis, in vitro synthesis by means of an expression system or by isolation from organisms. Proteins according to the invention can contain one or more amino acid analogues or amino acids not naturally occurring. Also, modifications (e.g., glycosylation, and the like) or labeling (e.g., biotinylation) according to the scientific knowledge can be contained.
The fragments preferably have a length of at least 3 to 5 amino acids, particularly preferred of 8 to 10 amino acids and in particular preferred of 11 to 15 amino acids. This is also true for the proteins according to the invention described below.
The lipoprotein OrfA according to the invention can be purified, for example, by a method that is based on the interaction of this protein with the PilC protein from Neisseria gonorrhoeae. It is preferably purified from homogenates of cells expressing this protein by means of chromatography matrices containing immobilized PilC protein. The protein can then be selectively eluted using its affinity to PilC and produced in essentially pure form.

The proteins according to the invention or fragments thereof can be used as immunogens for the production of antibodies. Therefore, the present invention also relates to antibodies that are directed against a protein according to the invention or a fragment thereof. The antibodies can be both polyclonal and monoclonal. Methods for the production of such antibodies are known to the skilled person.

In a preferred embodiment such antibodies are directed against epitopes of the protein according to the invention or fragments thereof that are important for the adherence and for the interaction with PilC.

The antibodies according to the invention can be, for example, produced by introducing the nucleic acid sequences according to the invention described above into hosts by in vivo transfection. Thereby, the protein or a fragment thereof is expressed in the host and the antibodies directed against them are induced (nucleic acid vaccination). This is also the case with the antibodies described below.

The present invention further relates to nucleic acid molecules having a length of at least 12 nucleotides and specifically hybridizing to the nucleic acid molecule

described above. Preferably, such nucleic acid molecules have a length of at least 15 nucleotides, particularly preferably of 20 nucleotides. Such molecules are, for example, suitable as primers for in vitro amplification, for example by polymerase chain reaction (PCR), or suitable for diagnostic purposes, that is for specifically identifying the nucleic acid molecules of the invention in samples.

The invention further relates to pharmaceutical compositions containing a nucleic acid molecule according to the invention described above, a protein, a biologically active fragment thereof and/or an antibody according to the invention described above. In the context of the present invention such pharmaceutical compositions can contain the usual pharmaceutical adjuvants, diluents, additives and/or carriers. The invention also relates to vaccines containing the nucleic acid molecules described above, proteins, biologically active fragments thereof and/or antibodies.

In a further aspect the present invention relates to diagnostic compositions containing the nucleic acid molecules according to the invention described above, proteins, biologically active fragments thereof and/or antibodies.

A further aspect of the present invention relates to receptors and substances having receptor function, interacting as ligands with the adhesin according to the invention, the OrfA-PilC complex. Such substances can be identified as competitive inhibitors of the adherence function due to their interaction with the OrfA-PilC complex. They can be surface components of human cells, particularly preferred surface components of human epithelial cells or chemical substances of any origin.

Finally, the present invention relates to inhibitors that influence the interaction between the OrfA-PilC adhesin complex and its receptors. Enclosed are all substances according to the invention that influence the interaction between the OrfA-PilC adhesin and its cellular receptor and therefore disturb the adherence. In a particularly preferred embodiment substances that irreversibly bind to the adhesin complex such as receptor analogues are encompassed.

Finally, the present invention relates to pharmaceutical compositions containing as an agent

(a) a receptor according to the invention;

(b) a receptor analogue according to the invention; and/or

(c) an inhibitor according to the invention,

optionally together with the usual pharmaceutical adjuvants, diluents, additives and carriers.

The pharmaceutical compositions described in the context of the present invention can be used for identifying and characterizing a bacterial sample not yet known as pathogenic Neisseria spc. and for diagnosing a Neisseria infection.

On the polynucleotide level, preferably hybridization probes are used containing the nucleotide sequences of the invention that are specific for one of the orf-gene regions or nucleotide sequences of the invention from one of the orf gene regions are used as primers for the PCR amplification of the genomic DNA region to be identified that is specific for pathogenic Neisseria.

On the polypeptide level diagnosis is preferably performed with the help of antibodies of the invention or, in the case of antibody screening tests, with the help of immunogenic proteins of the invention or fragments thereof.

Receptors, receptor analogous substances and inhibitors of the interaction between the OrfA of the invention and the corresponding receptors of the host cells can be used as therapeutics for infections at an early stage or if an infection is suspected. By strongly inhibiting the adherence, the adhesion of the pathogens to the epithelial host cells can be prevented so that by the usual defense mechanisms, such as ciliary movement of the epithelial cells, mucus secretion, mass flow of body fluids and the like, the pathogens can be eliminated.

Finally, the pharmaceutical compositions of the invention can be used for preventing or fighting Neisseria infections. Preferably, for preventive applications the proteins of the invention or fragments thereof are used for the production of a vaccine for active immunization, or antibodies of the invention are used for the production of a passive vaccine applicable as a therapeutic. The applications described above also apply to the pharmaceutical compositions and diagnostic compositions described below.

The subject matter of the invention further relates to nucleic acid molecules encoding a protein or a biologically active fragment thereof from bacteria of the genus Neisseria having the amino acid sequence described in Seq ID No. 3. Such nucleic acid molecules preferably have the nucleotide sequence described in Seq ID No. 3, in particular the one of the described coding region. The invention also relates to nucleic acid molecules the sequence of which deviates from the sequences of the molecules mentioned above due to the degeneration of the genetic code. Also nucleic acid molecules are the subject matter of the invention that hybridize to the nucleic acid molecules mentioned above (for the definition of the term "hybridization" see above). For the possible variants of the nucleic acid molecules the same is true what has already been described in connection with the nucleic acid molecules encoding OrfA.

The invention also relates to vectors containing the described nucleic acid molecules, in particular those in which they are linked to regulatory DNA elements for the expression in prokaryotic or eukaryotic cells, as well as to host cells that contain such vectors or that are genetically manipulated with the described nucleic acid molecules.

The invention also relates to proteins encoded by the nucleic acid molecules described above and to proteins containing amino acid sequences that immunologically cross-react with the amino acid sequence depicted in Seq ID No. 3 or fragments thereof. In the context of this invention they are called OrfI proteins. The protein from Neisseria gonorrhoeae having the amino acid sequence depicted in Seq ID No. 3 shows in the T7 expression system an apparent molecular weight of about 18 kd. A homology to presently known proteins could not be shown. Experimental data indicate that the protein is located intracellularly and possibly has a regulatory function.

This protein can be produced by a method in which a host cell described above is cultivated under conditions allowing the expression of the protein and in which the

protein is obtained from the cells and/or the culture supernatant. Therefore, the invention also relates to proteins obtainable by such a method.

The invention also relates to antibodies against a protein described above or a fragment thereof as well as to nucleic acid molecules having a length of at least 12 nucleotides and specifically hybridizing to a nucleic acid molecule described above. Preferably, the molecules have a length of more than 15 nucleotides and particularly preferably of more than 20 nucleotides.

The invention further relates to pharmaceutical compositions containing a nucleic acid molecule, protein, biologically active fragment thereof and/or an antibody described above and, optionally, a pharmaceutically acceptable carrier.

The invention further relates to diagnostic compositions containing the nucleic acid molecules, proteins, biologically active fragments thereof and/or antibodies described above.

The subject matter of the invention further relates to nucleic acid molecules encoding a protein or a biologically active fragment thereof from bacteria of the genus Neisseria that has the amino acid sequence depicted in Seq ID No. 4. Such nucleic acid molecules preferably have the nucleotide sequence depicted in Seq ID No. 4, in particular the one of the indicated coding region. The invention also relates to nucleic acid molecules the sequences of which deviate from the nucleotide sequence of the above-mentioned molecules due to the degeneration of the genetic code. Furthermore, the subject matter of the invention also relates to nucleic acid molecules hybridizing to the above-mentioned nucleic acid molecules (for the definition of the term "hybridization" see above). The same applies to possible variants of the nucleic acid molecules as has already been described in connection with the nucleic acid molecules encoding OrfA.

In a preferred embodiment the above-described nucleic acid molecules encode a protein that is able to form a complex with the protein PilC and therefore shows an ability of adherence to human cells.

The invention also relates to vectors containing the described nucleic acid molecules, in particular those in which they are linked to regulatory DNA elements for the expression in prokaryotic or eukaryotic cells, as well as to host cells that contain such vectors or that have been genetically manipulated with the above-described nucleic acid molecules.

The invention also relates to proteins encoded by the above-described nucleic acid molecules and to proteins containing the amino acid sequences that immunologically cross-react with the amino acid sequence depicted in Seq ID No. 4 or parts thereof. These are called OrfB in the context of the present invention. The protein from Neisseria gonorrhoeae having the amino acid sequence depicted in Seq ID No. 4 shows in the T7 expression system an apparent molecular weight of about 57 kd. A homology to presently known proteins could not be shown. Experimental data indicate that the protein is, like OrfA, located at the cell surface and is accessible from the outside. Furthermore, it obviously also possesses the ability to form a complex with the protein PilC and to induce either alone or in combination with OrfA the adhesion to human cells.

This protein can be produced by a method in which an above-described host cell is cultivated under conditions allowing the expression of the protein and in which the protein is obtained from the cells and/or the culture supernatant. Therefore, the invention also relates to proteins obtainable by such a method.

The invention also relates to antibodies against an above-described protein or fragment thereof, as well as to nucleic acid molecules having a length of at least 12 nucleotides and specifically hybridizing to an above-described nucleic acid molecule. Preferably, such molecules have a length of more than 15 nucleotides and particularly preferred of more than 20 nucleotides.

Furthermore, the invention relates to pharmaceutical compositions containing an above-described nucleic acid molecule, protein, biologically active fragment thereof and/or antibody and, optionally, pharmaceutically acceptable carriers.

The subject matter of the invention further relates to diagnostic compositions containing the above-described nucleic acid molecules, proteins, fragments thereof and/or antibodies.

Illustration of the figures and the sequence protocols:

Figure 1 schematically shows the construction of the plasmid pES25.

Figure 2 shows the nucleotide sequence (SEQ ID No. 1) of the orf gene region, starting from position 1 at the modified BglI cleavage site and ending with position 3260, the last nucleotide of the HindIII cleavage site. Restriction cleavage sites, ribosome binding sites (Shine-Dalgarno sequences) and promoter sequences (-35 and -10 regions) are labeled.

SEQ ID No. 1 further shows the amino acid sequences of the proteins OrfI, OrfA and OrfB encoded by the *orf* gene region. The amino acids of the lipoprotein signal sequence of OrfA are written in italic, the cleavage sites of the lipoprotein signal peptidase II is labeled with the tip of an arrow. The amino acid cysteine that represents the amino terminal of the processed OrfA lipoprotein and is modified to glyceryl cysteine with fatty acid is marked with a circle. The first seven amino acids of OrfB that are similar to a typeIV-pilin-signal sequence are written in bold. The labeling between amino acids 7 and 8 and between 11 and 12 characterize potential cleavage sites analogous to the processing of the typeIV-pilin.

Seq ID No. 2 shows the nucleotide sequence of the gene region encoding OrfA as well as flanking sequences. The amino acid sequence of OrfA is depicted, too.

Seq ID No. 3 shows the nucleotide sequence of the gene region encoding OrfI as well as flanking sequences. The amino acid sequence of OrfI is depicted, too.

Seq ID No. 4 shows the nucleotide sequence of the gene region encoding OrfB as well as flanking sequences. The amino acid sequence of OrfB is depicted, too.

The examples illustrate the invention.

Examples

Example 1

Method for the isolation of the lipoprotein adhesin OrfA:

During the chromatographic purification of the PilC protein a decisive observation with regard to the identification of the new adhesin of Neisseria gonorrhoeae of the invention was made. A recombinant PilC protein was used that was amplified by an oligo-histidine region with six histidine residues (His_6-tag) in order to make the chromatographic purification easier (Rudel et al., Nature 373, 357-359, 1995). The amplification of the protein by the histidine hexapeptide makes the selective binding to a nickel-nitrilotriacetate-agarose matrix (Ni-NTA matrix) possible. After the cell wall fraction produced from cultures of a pilus-free PilC overexpression strain N560 (Rudel et al., see above) from Neisseria gonorrhoeae had been extracted, the extract was loaded on an Ni-NTA chromatography matrix. Usually, for the method that was developed for the purification of recombinant PilC unspecifically bound material was removed by extensive washing with a buffer containing imidazole. However, in the first elution fraction a protein of 36 kd (OrfA) could be identified together with PilC in an approximately equimolar ratio.

For the preparation of the PilC-OrfA protein fraction the strain N560 from Neisseria gonorrhoeae was plated on 30 GC-agar plates and incubated in 5% CO_2 at 37 °C for 20 hours. The GC-agar medium (GC agar base, Becton Dickinson, Heidelberg) contained the usual additional factors necessary for the growth of Neisseria gonorrhoeae (0.1 mg vitamin B12, 10 mg adenine, 0.3 mg guanine, 100 mg

glutamine, 1 mg cocarboxylase, 0.3 mg thiamine, 259 mg L-cysteine, 11 mg L-cystine, 1.5 mg arginine, 5 mg uracil, 0.2 mg $Fe(NO_3)_3$, 2.5 mg nicotineamide-adenine dinucleotide, 0.13 mg p-aminobenzoic acid and 1 g dextrose per 1 liter of medium) that were added as a sterile filtrate to the GC basis medium after heat sterilization. Furthermore, the so supplemented GC agar medium contained 5 µg/ml tetracycline and 100 µM IPTG. The bacterial lawns were removed with cotton pads, transferred to 30 ml of washing buffer (Tris-HCl pH 8.0 with 0.15 M NaCl) and centrifuged at 4,000 rpm, 4 °C for 15 minutes (Du Pont Sorvall Centrifuge RC-5B, Rotor SS-34). The cell sediment was again resuspended in 30 ml of washing buffer, and the bacteria were broken up by ultrasonic homogenization after lysozyme and 5 mM EDTA Na_2 had been added. Intact bacteria were separated by centrifugation at 5,000 rpm at 4 °C for 15 minutes. The cell coats of the lysed bacteria were sedimented by centrifugation of the supernatant at 20,000 rpm at 4 °C for 60 minutes and taken up in 10 ml of washing buffer additionally containing 10 % glycerine, 10 mM $MgCl_2$ and 2 % Triton X-100. After an incubation of 45 minutes at 37 °C they were centrifuged again (20,000 rpm, 4 °C for 60 minutes) and the membrane sediment suspended in 10 ml of washing buffer with 10 % glycerine, 10 mM $MgCl_2$ and 2 % LDAO (N,N-Dimethyldodecylamin-N-oxide) and incubated at 37 °C for 60 minutes. After they were centrifuged again (20,000 rpm, 4 °C for 60 minutes), the supernatant containing the biologically active PilC-OrfA complex was subjected to a nickel-chelate-affinity chromatography for further purification. For this purpose a Ni-NTA-gel matrix (300 ml bed volume) was washed with 5 bed volumes of aqua bidest. and loaded with 10 ml of the supernatant. Unspecifically bound proteins were removed by elution with 5 column volumes of 50 mM imidazole in PBS buffer pH 8.0. After the column had been washed again with 5 to 10 bed volumes 20 mM sodium phosphate pH 7.5 with 0.15 M NaCl (PBS buffer) the biologically active PilC-OrfA complex was eluted with a citrate/phosphate buffer (10 mM citric acid, 1 M sodium phosphate, pH 3.5, 10 % glycerin, 0.15 M NaCl) in the first elution fraction and instantly neutralized with a 1 M Na_2HPO_4 solution. The eluate containing PilC and OrfA was frozen in liquid nitrogen and strored at -70 °C.

Example 2

Isolation of the polynucleotide sequence carrying the orf-gene region

To further characterize the 36 kd OrfA protein, mice were immunized with the PilC-36 kd protein fraction. The 36 kd protein proved to be very immunogenic. With the antibodies obtained this way a pBA plasmid gene library of the Neisseria gonorrhoeae MS11 genome in E. coli GC1 was screened for the presence of antigens. Several clones showing a positive reaction were isolated and clone H1967 was chosen for further characterization.

The library plasmid pES25 (Figure 1) of clone H1967 contained a genomic fragment of approximately 11 kb, cloned in vector pBA. Restriction fragments of the total region were subcloned in pUC and pBluescript KS (+) vectors, respectively. On the basis of the expression of the derived plasmids in minicells and immunoblotting analyses subclones were chosen producing the 36 kd protein. The subclones were used for sequencing. The sequences were determined by directly sequencing restriction fragments, by sequencing continuously shortened ExoIII nuclease fragments of the BglII-PstI fragment (positions 1 to 2560 of Seq ID No. 1), as well as by sequencing PCR amplified fragments.

The region depicted in SEQ ID No. 1 starting from the BglI cleavage site (position 1) to the HindIII cleavage site (position 3260) had three open reading frames with a high coding probability with each reading frame beginning with the start codon ATG, having a ribosome binding site that precedes the start codon in a suitable distance (S.D. sequence) and ending with a stop codon.

The three reading frames have the same orientation. The first open reading frame starts at position 136 of the sequence depicted in SEQ ID No. 1 and ends at position 450 with the stop codon TAA. The encoded protein was called OrfI and had an apparent molecular weight of 18 kd in the T7 expression system.

No significant homologues could be identified by sequence comparison in the EMBL gene library (Release 43.0 from 6/95) and in the SwissProt data bank (Release 31.0

from 3/95), neither on a nucleotide sequence level nor on an amino acid sequence level.

The second open reading frame starts at position 583 and ends at position 1545 with the stop codon TGA. It encodes the OrfA protein having an apparent molecular weight of 36 kd in the T7 expression system. Also to this sequence no significant homologues could be detected via data base search. The sequence analysis by means of the protein analysis program "Motifs" (GCG Genetics Computer Group, Inc., Madison, Wisconsin, USA) showed, however, a complete homology of the N-terminus of OrfA to lipoprotein specific signal sequences (position 583 to 636). The characterization of OrfA as a lipoprotein could be substantiated by experiments (vide infra).

The third open reading frame starts at position 1585 and ends at position 3114 with the stop codon TGA. The protein OrfB hereby encoded has an apparent molecular weight of 57 kd in the T7 expression system. Also to this reading frame no homologue could be identified via data base search.

As a structural peculiarity the amino terminus of the OrfB sequence displays a signal sequence showing similarities to the type IV-prepilin signal sequence. At positions 8 and 12 of the amino acid sequence there is phenylalanine so that there are in addition two possible cleavage sites for the type IV pilin signal peptidase. It can be derived herefrom that OrfB presumably is a secreted protein.

The molecular weights of all the three gene products measured in the T7 expression system correspond to the values theoretically calculated from the sequence. The separation of the expression products by means of gel electrophoresis showed that the OrfB-band was significantly weaker than the OrfA-band in all the cases. This points to a weaker expression of OrfB.

Two regions showing a sequence homology to the promoter regions were identified. One of them is located in front of the orfl gene, the second one in front of the orfA

gene, each leaving an appropriate distance (SEQ ID No. 1). Therefore, it can be assumed that orfA and orfB form a transcription unity.

The analysis of the Neisseria gonorrhoeae MS11 genome after ClaI and MluI digestion showed a complex band pattern in Southern hybridization with plasmid pES-8 as sample. This fact indicates the existence of several copies of the orf-gene region, probably of three copies, in the genome of Neisseria gonorrhoeae MS11. If all these loci are expressed, if they are subjected to antigenic variations like, for example, the Neisseria genes pilS and opa, and if the flanking regions of the orf gene region are involved in the sequence repetitions, is presently not known.

Example 3

Characterization of the localization of OrfA and OrfB on the cell surface

In order to experimentally prove the lipoprotein nature of orfA derivable from the perfect structure homology of the amino terminus of *orfA* to lipoprotein signal sequences, both N. gonorrhoeae and E. coli recombinants transformed with the orf-gene region were labeled with [^{3}H] palmitate. The results of the labeling show that in all the cases, both with N. gonorrhoeae and with the E. coli recombinants, lipoproteins in the corresponding molecular weight range could be identified. While with N. gonorrhoeae several proteins were labeled and the labeled band could not be precisely assigned since there was no *orfA*$^-$ mutant available, the *orfA* recombinants of E. coli showed in comparison to the control strain unambiguously only one additional band having the molecular weight of OrfA. An OrfA fusion protein that was tested in addition and was amplified at the carboxy-terminal by a fusion of 3 kd, also had a [^{3}H] palmitate labeling and migrated to a position precisely corresponding to the molecular weight that was, as expected, increased due to the fusion.

When prepared cell coats were treated with detergents, OrfA showed a solubility that is typical of proteins of the outer membrane. By separating the cell coat by means of density-gradient centrifugation it could be confirmed by means of marker proteins that OrfA was located in the outer membrane of N. gonorrhoeae. Also with orf

recombinants of E. coli, OrfA was shown to be a protein component of the outer membrane by means of said method.

The accessibility of the cell surface was proven by means of an immunofluorescence test both for OrfA and OrfB. A defective pilC mutant of Neisseria gonorrhoeae the two pilC genes of which were switched off is labeled by the PilC-OrfA antiserum in the same way as recombinant E. coli strains carrying the orf-gene region. The non-transformed control strain showed, as was to be expected, a negative reaction. A positive reaction in the immunofluorescence test of N. gonorrhoeae and orf recombinant E. coli strains could be brought about by means of OrfA and OrfB specific antisera using purified fusion proteins of either OrfA or OrfB for the production of these antisera. If antisera were used directed against an Orfl fusion protein, the immunofluorescence test with N. gonorrhoeae was negative. From this it can be deduced that OrfA and OrfB are located on the cell surface and are accessible from the outside, whereas Orfl probably is located intracellularly.

The surface localization of OrfA and OrfB could only be proven in recombinant E. coli strains carrying the whole *orf* region.

Example 4

Adhesin property of the OrfA-PilC complex

As mentioned above OrfA could be obtained in pure form by chromatography on an Ni-NTA-chelate matrix due to its affinity to PilC. Since the function of PilC as pilus associated adhesin had been proven and the direct binding of PilC to human ME-180 cells had been known, it was obvious to test the adherence property of the PilC-OrfA complex. The experiments were performed with the E. coli strain HB101 (E141) since it does not possess the mannose specific typel pili and shows almost no binding to human ME-180 and Chang epithelial cells. After the transformation of HB101 with the plasmid pES25, no adherence, neither to ME-180 nor to Chang cells, could be mediated. If the same recombinants, however, were pre-incubated with PilC

protein, a strong adherence to Chang epithelial cells but not to ME-180 cells could be induced (Table I).

Table I

OrfA-dependent modulation of the PilC mediated adhesin function

	Adherence to human epithelial cells	
	ME180 cells	Chang cells
N. gonorrhoeae, Orf+ PilC+, Pili+	+++	+
N. gonorrhoeae, Orf+, PilC+, Pili-	+	+++
E. coli (E141)	-	-
E. coli (E141) + PilC (extern)	-	-
E. coli (H2561)	-	-
E. coli (H2561) + PilC (extern)	-	-
E. coli (H2560)	-	+
E. coli (H2560) + PilC (extern)	-	+++

Three independent experiments were evaluated, whereby the adherence of Neisseria was determined using 500 cells and the adherence of the E. coli strains was determined per epithelial cell.

+++ 100 %, ++ 50 %, + 25 % adherence.

E. coli E141 = E. coli strain HB101 without plasmid; E. coli H2561 = E. coli strain HB101 with plasmid pBA; E. coli H2560 = E. coli strain HB101 with plasmid pES25

The plasmid pES25 (Figure 1) is a pBA vector containing a genomic fragment from Neisseria gonorrhoeae of approximately 11 kb carrying the coding regions *orfA*, *orfB* and *orfI*.

The E. coli strain H2560 was deposited at the Deutsche Sammlung für Mikroorganismen (DSM, Braunschweig, Germany) under the DSM-Accession Number DSM 10257.

The result obtained is surprising since pilus carrying Neisseria bind to ME-180 cells with a significantly higher affinity than to Chang epithelial cells. This result can be put down to the fact that PilC has different adherence properties depending on its localization. As an adhesin component in the pilus PilC preferably binds to receptors of the ME-1800 cell surface, whereas as an adhesin located on the cell surface in the complex with OrfA PilC preferably recognizes receptors on Chang epithelial cells. If in the latter case adhesin properties also can be ascribed to OrfA and/or OrfB, is presently not known.

The results obtained for recombinant E. coli strains could be reproduced with the same result with N. gonorrhoea. If the pilus-free strain N 300 (P- Opa-), which hardly binds to ME-180 or Chang cells, is pre-incubated with purified PilC, the adherence to Chang epithelial cells can be significantly increased.

The described experimental approaches obviously provide for a model that is suitable to analyze a mechanism for the modulation of the adherence properties, how they can in cascade-like order effect the increasingly strong adherence of the pathogens to the host cells or how they can be the basis for the tissue tropism.